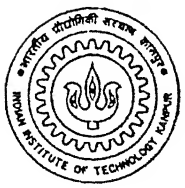
# An Image Processing Approach to Differential Blood Count

*A Thesis Submitted*
*in Partial Fulfillment of the Requirements*
*for the Degree of*
*Master of Technology*

*by*

**Nishit Desai**




*to the*

**Department of Computer Science & Engineering**
Indian Institute of Technology, Kanpur

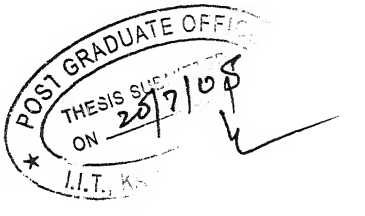
**July, 2005**

# Certificate

This is to certify that the work contained in the thesis entitled "*An Image Processing Approach to Differential Blood Count*", by *Nishit Desai*, has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

July, 2005

(Dr. Harish Karnick)
Department of Computer Science & Engineering,
Indian Institute of Technology,
Kanpur.



## Abstract

Medical image analysis plays a major role in providing quality health care. Better imaging techniques have enabled effective diagnosis of various diseases. Automation of the diagnostic processes is essential because the manual methods require considerable amount of time, effort and care, besides being prone to errors. It also facilitates testdata collection for the people in remote areas, where expert physicians are not widely available.

This thesis presents a step towards testing of peripheral blood. Our works aims at developing an automatic and reliable system for obtaining the Differential blood count (DLC), an important measure for blood-related diseases. We present an effective technique for identification and segmentation of white blood cells in smear images using autothresholding and the watershed algorithm. Feature extraction and classification based on the shape-color and the texture based feature was performed with peak accuracy of 80%. The performance of the system was satisfactory considering the scarcity of data and relatively poor quality of images.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

## 1.1 Medical Image Analysis

Aim of medical image analysis is to develop systems that are capable of processing medical images required for diagnosis. The primary purpose of medical image analysis is to extract relevant information from images in order to facilitate unambiguous detection of abnormalities. Such systems can also be used for visualization. Computer-based descriptions are often more consistent than those derived by human observers. The descriptions can include shape, color, pattern, texture, and other image features. The increasing availability of computing power and appropriate modeling techniques have enabled rapid development of medical systems for quantitative image analysis that support disease detection, therapy planning and medical education.

A wide range of tasks in medical image analysis include image enhancement, segmentation, noise removal, pattern detection depending upon the requirement. For example, MRI image segmentation[9] helps in diagnosis, but sometimes the low-contrast MRI images need enhancement before being used for diagnosis[34], while the ultrasound fetal images would need to be analyzed for textures in order to determine lung maturity[5]. Patterns in bone images are recognized for forensic applications as well as for the determination of age, study of age-related bone developments and

one diseases[27]. Angiogram, the X-ray image of network blood vessels needs to be rocessed for suppressing the shadows created by bones, in order to enable correct liagnosis[38]. The images of skin moles need processing to extract descriptions that an aid the diagnosis of melanoma, a skin cancer[17]. Similarly, segmentation of putum images helps in the diagnosis of lung cancer[35]. Range of diseases caused by disorders in blood[29] can be diagnosed with the help of blood smears. In this hesis, we have worked with color images of blood smears to detect and classify white blood cells.

## 1.2 Composition of Blood

Blood is a fluid tissue flowing through the circulatory system transporting the digested food substances, excretory products, and dissolved gases. It is composed of various types of blood cells suspended in a fluid called plasma. The types of blood cells include,

- Red Blood Cells (RBC) or Erythrocytes, which carry oxygen from the lungs to the rest of the body.
- White Blood Cells (WBC) or Leukocytes, which help fight infections and aid in the immune process.
- Platelets or Thrombocytes, which help in blood clotting.

Blood cells[15] are formed in the bone marrow. In the initial phase, they are called "stem cells" or "hematopoietic cells". As the stem cells mature, distinct cells of each type evolve.

WBC's are responsible for the defense system in the body by fighting infections. They are much bigger in size and fewer in number than RBC's. There are approximately 6,000 WBC's per cubic millimeter of blood or half a million WBC's in every drop of human blood. The WBC's have a life-span ranging from few hours to few days. When they die, the dead ones are engulfed by the surrounding WBC's and the dead cells are replaced with new ones. WBC's include immature and mature types.

Studies show that WBC's display features that continuously evolve from the primitive forms to the mature cell types, making the initial and final features extremely different. This feature variability makes WBC classification a difficult task. Immature types include unsegmented neutrophils, blasts, variant lymphocytes, proerythroblast, myeloblast, erythroblast and monoblast. Mature cells can be divided into 5 classes, namely: neutrophils(50–70%), lymphocytes(25–35%), monocytes (4–10%), eosinophils(less than 5%), basophils(fewer than 1%).

***Complete Blood Count:*** Complete blood cell count is the measurement of size, number, and maturity of the different blood cells in a specific volume of blood, usually a microliter. This is used to determine abnormalities with either the production or destruction of blood cells. Variations from the normal number, size, or maturity of the blood cells is an indication of infection or disease, like leukemia, anemia and sickle cell disease. One of the steps in 'Complete Blood Count' is to perform 'Differential Blood Count'.

***Differential Blood Count:***Differential blood count is specific to WBC's. It it carried out to calculate the relative percentage of each type of WBC, since it helps in diagnosing the cause of many ailments. In a normal person, there are about 3150 to 6200 neutrophils, 1500 to 3000 lymphocytes, 300 to 500 monocytes, 50 to 250 eosinophils and 15 to 50 basophils, per microliter of blood. Changes in these counts are indicators of a disease. For instance, a high neutrophil count would suggest infection or cancer or physical stress. High lymphocyte counts are usually due to Acquired Immune Deficiency Syndrome (AIDS). High monocyte and eosinophil count usually point at bacterial infection. Thus differential blood count is indicative of one's health status. Both conventional and automatic methods are prevalent in computing it accurately.

## 1.3 Motivation

Manual method for differential blood count uses a stained slide. The technician typically studies 100 WBC's to determine the type of each of the cells, in order to

calculate their relative percentages. This method suffers from several pitfalls. It's not only time-consuming but the quality of the results highly depends on the technician's skill and experience. Manual analysis is questionable because of precision and the poor reproducibility of the results apart from the amount of work involved.

Some automatic methods utilize fluid properties of blood for counting purpose. WBC counters based on flow cytometry[1] are in use. They utilize the Coulter Principle[12] of impedance measurement for a liquid-dispersed blood flow. The distribution of RBC's, WBC's and the classification of a couple of classes of WBC's are performed using laser light scattering from stationary suspensions[36]. These methods rely on haemetological practice, but at the same time they ignore rich visual information available in image. Hence image processing offers a better alternative for the task.

Automated image processing based robust systems can overcome manual errors. Besides, Automated systems can help overcome scarcity of trained personnel. Given a full-fledge system for differential blood count, the only manual effort required would be to prepare a stained slide, acquire images with a microscope and hand it over to the system, which can easily be done by suitably training a person.

## 1.4 Organization of Thesis

Chapter 2 describes the typical steps in automated WBC counting and summarizes the past work for automatic and semi-automatic systems for differential blood counting based on image processing.

Chapter 3 presents the segmentation scheme employed for the system. The technique based on $k$-means clustering followed by autothresholding and the watershed algorithm is proposed and results of the same are presented.

Chapter 4 discusses various shape based, color based and texture based features from the cell images in order to facilitate the feature extraction process. Along with

the features finally chosen to form the feature vector and the feature rejected are also discussed.

Chapter 5 presents the comparative study of classification task and presents results for the same. The relative performance of the classifiers is tabulated.

Chapter 6 contains the concluding remarks and possible directions for the future work.

# Chapter 2

# Overview and Previous Work

## .1 Overview

system for automatic differential blood count aims to distinguish between the five asses of mature WBC's, namely lymphocyte, monocyte, eosinophil, basophil and eutrophil. Typically, the input to the system is a digital image of blood smears nd the output is the differential count of the type of cells.

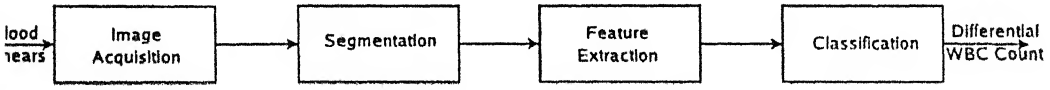



Figure 2.1: Block Diagram for typical Differential Blood Count System

The block diagram of the overall system is depicted in Figure 2.1. The main ages of system are :

***cquisition:*** This is the process of capturing images of blood smears. In our se, color images are captured using a digital camera mounted on a microscope.

***egmentation:*** This is a very crucial step, because this step estimates the shape WBC and thus efficiency of the subsequent stages depends on this. In this stage, /BC's are extracted from the background and other constituents of blood such as

RBC's, plasma, platelets and cell-fragments. Further, distinction between nucleus and cytoplasm for each cell is accomplished.

***Feature Extraction:*** Features extracted from segmented cells are generally shape-based, color-based and texture-based. Shape-based features include eccentricity of the nucleus, eccentricity of the cytoplasm, area-ratio between the cytoplasm and the nucleus, number of nucleus lobes etc. Color-based features used are average red, blue, green component for nucleus and cytoplasm. While various texture-based features are energy, entropy, correlation etc. It has been observed that shape-based features are most important features.

***Classification:*** Based on the features extracted, a set of feature vectors, representing all 5 types of WBC's is created. This set is used to train the classification model. After training, an unseen sample of WBC can be classified as one of the 5 types. For classification purpose, widely used classifiers are neural networks, support vector machines and bayesian classifier.

## 2.2 Previous work

Previous work in Automated Differential blood count has been done at various levels. Some techniques have only worked on detecting WBC's in the image [39, 41], while some others have been successful in segmentation and classification of blood cells also [24, 25, 30]. Most of the work has been carried out on healthy and mature cells, but a couple of works [3, 30] also include immature cells for classification. Some of the techniques have used gray images as input [6, 39], while recent work has been done on color images [24, 25]. The important advantage of gray scale images is that they are less sensitive to variations of lighting conditions and require less processing time and storage as compared to color images, but at the same time important color information which could be useful for segmentation and classification remains unexploited. A review of some of the earlier work is presented.

## 2.1 Identification of WBC's

ssan Sheikh *et al.* [39] presented a method to differentiate between RBC's, WBC's l platelets from gray images on the basis of size, shape, volume of the cells and sence of nucleus. Cells were manually segmented from the image, followed by velet based feature extraction. Finally, an ALOPEX neural network [43] was used classification. Training and testing was done on 11 and 9 cell images respectively, ich is a very small database, but they claim an accuracy of 89%. Park *et al.* ] suggest a method based on gray scale bone marrow images for distinguishing erent type of cells. The technique uses the watershed algorithm to perform rsegmentation to create initial "patches" and patch labels are adjusted using ntext information till convergence. However, they do not specify any objective luation for effectiveness of the technique. Sobrevilla *et al.* [41] have reported omatic WBC detection in gray scale bone marrow images. Geometrical, textural l morphological information are used to make fuzzy rules in order to detect the BC's. Accuracy of 93% has been reported for detection of WBC's. Wei *et al.* ] used neural networks to distinguish between the RBC's, WBC's and platelets. ey proposed a "boundary following" technique to retain details of contour of cells. ey used scanned images from *Atlas of Blood Cells* [47] for their input. They claim )% accuracy when boundary detection is accurate, however the technique is not ally automatic. It expects threshold values for differentiation of cells and starting nt for boundary detection from the user.

Hengen *et al.* [21] have reported a technique to follow the identification stage. ey presented a declustering method to handle overlapping cells. Thresholding the tance transform, followed by a region-growing algorithm is used for such cells. mford *et al.* [2] have achieved accurate nucleus segmentation for pap-stained vical images, which can be extended for blood cell segmentation.

## 2.2 Segmentation of WBC's

maniciu *et al.* [10] use non-Gaussian cluster in $L * u * v$ color space. Their cell mentation algorithm detects clusters in $L * u * v$ color space and refines their

border using gradient ascent mean shift procedure [11]. Katz [24] extracts region of interest based on thresholding. The segmentation of subimage into cell and non-cell regions is carried out using Canny Edge Detection followed by circle identification. However, the threshold was selected empirically and and circle identification required manual intervention. Wermser *et al.* [45] used Hierarchical Thresholding based on chromatic properties of background and cell components. Kovalev *et al.* [25] uses a three-step algorithm of extraction of nucleus, circle-shaped approximation and improvement in cytoplasm region using *a priori* information. Cseke [13] implemented a fast segmentation technique which utilised Otsu's Automatic Thresholding Method [31]. However, they do not differentiate between red blood cells and cytoplasm.

### 2.2.3 Classification of WBC's

Bikhet *et al.* [6] have worked on segmentation and classification of 5 types of WBC's. Segmentation was achieved using Hierarchical Thresholding based on Histogram Entropy Classification and Iterative Threshold Selection [22, 33]. It is claimed that the 10-dimensional feature vector, shape-based and color-based, achieved accuracy of 90% on 71 cells. Ongun *et al.* [30] worked on color images containing both mature and immature cells. Segmentation was accomplished by morphological preprocessing combined with fuzzy patch labeling. 57-dimensions consisted of shape-based (area of nucleus and cytoplasm, ratio of nucleus to cell area etc.), color-based(Color historgram, mean and standard deviation of components in CIE-Lab domain), texture-based(contrast, homogeneity, entropy derived from the gray-level co-occurence matrix) features. While various classifiers have been used, peak performance has been achieved with 91% using SVM. Sinha [40] has worked on very good quality, colour images for 5 types of mature cells. Segmentation was done in two parts : Coarse segmentation was achieved using $k$-means, followed by fine segmentation using Expectation-Maximization Algorithm and achieved 80% segmentation accuracy. Features extracted were eccentricity of the nucleus and the cytoplasm, compactness of the nucleus, number of nucleus lobes, average red, green, blue components for the nucleus and the cytoplasm and texture-based feature namely energy, entropy and correlation. For classification various classifiers like nearest neighbour,

k-nearest neighbout, neighted $k$-nearest neighbour, bayesian classifier, SVM and neural networks have been used with SVM giving peak performance of 94%.

# Chapter 3

# Segmentation

## 3.1 Introduction

A typical blood smear consists of WBC's, RBC's, plasma, platelets and cell fragments. The goal of segmentation is to locate the WBC's in the smear image and mark the boundaries of nucleus and cytoplasm regions. This is necessary before further processing to classify them as one of the 5 classes of WBC's. This stage is very crucial because accuracy of classification will largely depend on outcome of segmentation.

Most techniques proposed so far are sensitive to the right selection of some parameters such as, image acquisition conditions [40], threshold selection [24], initial contour [30]. Also some of the techniques [24] assume circular shape for white blood cells, which is not true in most cases. While Cseke *et al.*[13] proposed a robust technique for segmentation, but further segmentation between cytoplasm and red blood cells was not attempted. We report a two-stage segmentation scheme that enables us to distinguish the cytoplasm and nucleus of WBC from the input image of a blood smear and requires no manual interaction for parameter tuning.

## 3.2 Segmentation

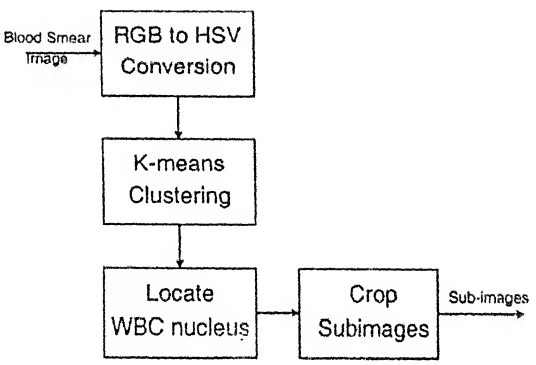
### 3.2.1 Overview

The figure 3.1 shows an overview of the proposed segmentation scheme. We first locate the nuclei of the cells using $k$-means clustering on the Hue-Saturation-Value(HSV) equivalent of the image. We then crop a rectangular region around each nucleus such that it encompasses the entire cell. The block diagram of the process for obtaining smaller images with only one cell, is shown in figure 3.1(a). Further processing is carried out on a gray version of these sub-images. Autothresholding [13], followed by declustering using the watershed algorithm is used to obtain segmented cytoplasm and nucleus regions. Results are further refined by choosing clusters that belong to the cytoplasm. The schematic for the second level of segmentation is shown in figure 3.1(b).

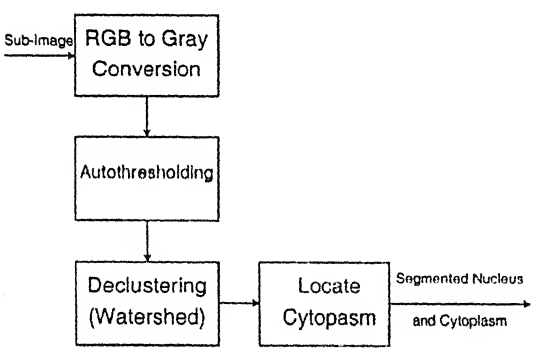
### 3.2.2 HSV-Space

HSV-space [16] is considered to be important for segmentation algorithms. Conversion to HSV in one approach to decouple the intensity component (Value) from the color information (Hue). This closely corresponds to the way color is perceived, rather than as superposition of the primary colors as in the RGB model. The HSV model is also useful for quantifying the purity of the color (Saturation).

A major difficulty with using color cues in machine vision is the color constancy problem which arises due to variation in color values brought about by lighting changes. This is particularly apparent in RGB space. Intensity is distributed through-out all three parameters, rendering color values highly sensitive to scene brightness. A simple approach to color constancy is to use the HSV color space which consists of hue angle (H), color saturation(S) and brightness(V). In order to obtain a limited level of intensity invariance, color can be modeled in HS-space. Hence this color space is generally preferred for segmentation algorithms.




(a) Generation of sub-images containing single cells




(b) Segmentation of nucleus and cytoplasm

Figure 3.1: Overview of the Segmentation

Our approach also performs segmentation in HSV space. The RGB images are converted to their HSV equivalent using the following equations:

$$H = \cos^{-1}\left[\frac{\frac{1}{2}\left[(R-G)+(R-B)\right]}{\left[(R-G)^2+(R-B)(G-B)\right]^{\frac{1}{2}}}\right] \quad (1)$$

$$S = 1-\frac{3}{R+G+B}min(R,G,B) \quad (2)$$

$$V = \frac{1}{3}(R+G+B) \quad (3)$$

Figure 3.2 shows a histogram of S-component of a typical cell, which depicts distinct peaks corresponding to each of the regions in the blood-smear, in which high values of saturation correspond to the WBC-nucleus. This feature helps us in identifying the cluster belonging to the nucleus.

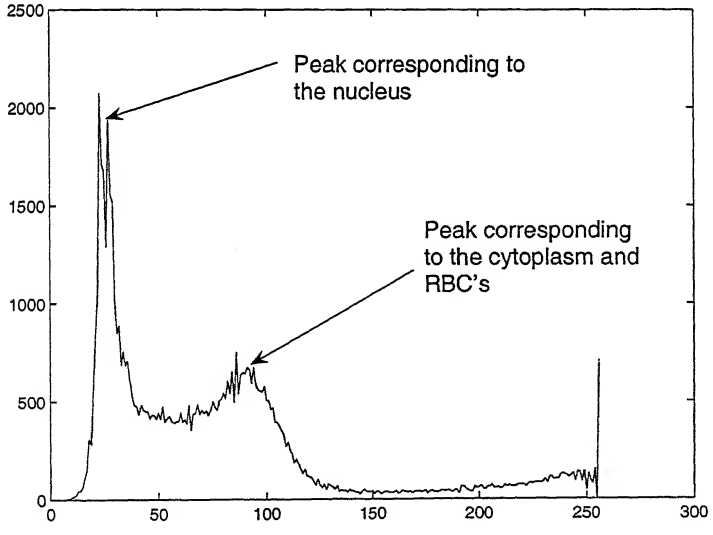



Figure 3.2: Histogram of S-component of a typical cell image

S, V. Thus, an image is a set of 3-component vectors in HSV space. For clustering purpose, all input values are normalized between 0 and 1. Smoothening is then performed on the image using a low-pass filter, by averaging over a windows size of $5X5$. $k$-means clustering is performed on this set of vectors. As mentioned in [40], we have used 5 clusters in our experiment, because apart from clusters for nucleus, cytoplasm, RBC's and background, the rim of the RBC's also forms a separate cluster. The centroids are initialized by choosing $k$ uniformly distributed points in the vector space. Euclidean distance is used as a measure of dissimilarity. When the difference between successive values of each centroid is less than a pre-defined threshold, clustering is said to have converged. At the end of clustering, each pixel is a member of one of the $k$ clusters and centroid for each cluster is obtained.

Among these clusters, we can say that the centroid with maximum saturation corresponds to the nucleus cluster. We then crop a rectangular region around nucleus of sufficient area such that it contains the entire cell. Thus a set of sub-images, containing a single WBC are obtained. Further segmentation is achieved using Autothresholding and Declustering using the watershed algorithm.

### 3.2.4 Autothresholding

Each sub-image is separately processed for this stage. It has been observed that given the gray scale image of any cell, dark regions correspond to the nucleus, bright regions correspond to background and intermediate regions correspond to cytoplasm and RBC's. So, thresholding the image into three classes, separate cell structures from one another, except cytoplasm and RBC's. We use automatic threshold selection proposed by Otsu [31]. In this method, optimal thresholds T1 and T2 are selected by maximizing interclass variance between dark, gray and bright regions. Cseke *et al.* [14] proved that maximizing interclass variance can be reduced to maximizing the function in equation 4.

$$E(T1,T2) = \frac{m^2(0,T1)}{n(0,T1)} + \frac{m^2(T1,T2)}{n(T1,T2)} + \frac{m^2(T2,L)}{n(T2,L)} \tag{4}$$

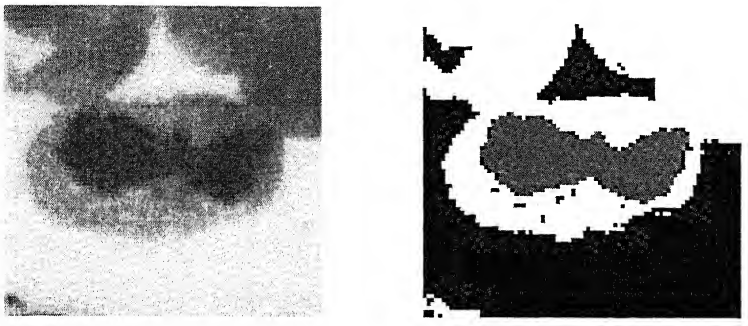


Figure 3.3: Original Image and Segmented Image

$$E(T1,T2) = \frac{m^2(0,T1)}{n(0,T1)} + \frac{m^2(T1,T2)}{n(T1,T2)} + \frac{m^2(T2,L)}{n(T2,L)} \tag{4}$$

where, L denotes number of gray levels (255 in our case) and $m()$ and $n()$ denote following expressions :

$$m(x,y) = \sum_{i=x}^{y-1} i \cdot H[i], \quad n(x,y) = \sum_{i=x}^{y-1} H[i], \quad y > x \tag{5}$$

where, $H[]$ denotes histogram of sub-image to be thresholded.

Further, Reddi *et al.* [37] proved that function E is maximized when equations 6,7 are satisfied, which can be easily solved using iterative algorithms.

$$\frac{m(0,T1)}{n(0,T1)} + \frac{m(T1,T2)}{n(T1,T2)} = 2 \cdot T1 \tag{6}$$

$$\frac{m(T1,T2)}{n(T1,T2)} + \frac{m(T2,L)}{n(T2,L)} = 2 \cdot T2 \tag{7}$$

Using threshold $T1$ and $T2$, we achieve coarse segmentation (see figure 3.3). As mentioned, using this autothresholding technique, we label pixel as belonging to either cytoplasm or RBC's. Declustering is done to separate cytoplasm from RBC's.

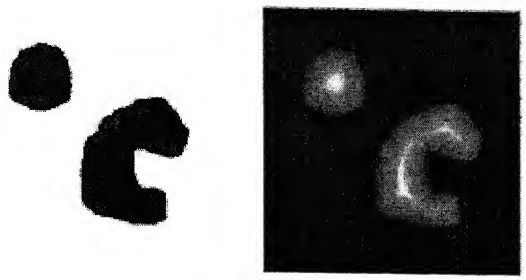


Figure 3.4: Negative Distance Transform

### 3.2.5 Declustering using watershed algorithm

Declustering technique involves concept of negative distance transform and the watershed algorithm.

#### Negative Distance Transform

The concept of distance transform [22] has been defined for binary images. It is computed for every foreground pixel, as the distance between the foreground pixel considered and the nearest background pixel. The result of the transform is a gray level image that looks similar to the input image except that the gray intensity of the points in the foreground region are changed to show the distance to the closest background pixel (see figure 3.4). The distance metric chosen is Euclidean, however other metrics can also be adopted. Distance transform is defined as:

$$D(f) = \{p : p = \min(f, b); \forall b \in B\} \tag{8}$$

where, $f$ is the foreground pixel and $B$ is the set of all background pixels in the image. Negative of this metric is defined as the Negative Distance transform.

$$N(f) = 255 - D(f) \tag{9}$$

#### Watershed algorithm

The watershed algorithm [4, 23] is a fundamental image segmentation tool in mathematical morphology. The watershed transform is based on an analogy with topographic reliefs. An image can be thought of as a three dimensional relief with the grayscale value at each point corresponding to height. Imagine that the relief has

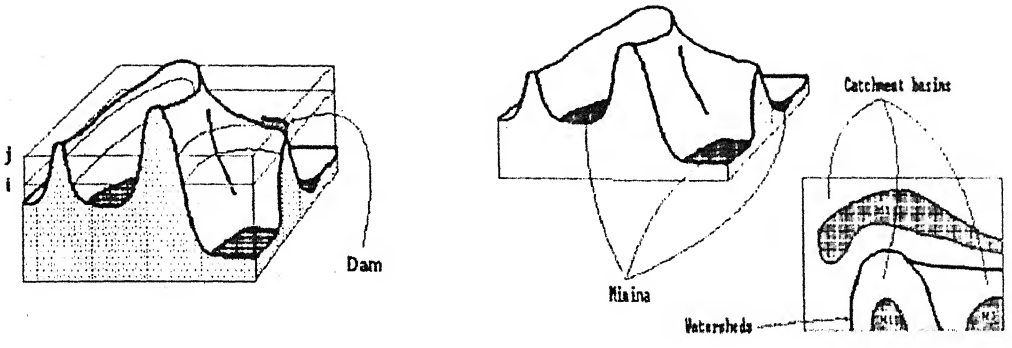



Figure 3.5: Visualization of Watershed Algorithm

a point. Once the relief has become completely covered by water, we end up with a structure with several barriers or dams on it. These dams represent the watershed lines and serve to separate the "catchment basins" of the relief (figure 3.5). One of the main advantages of the watershed transform as a segmentation tool is that the segment boundaries it produces are closed.

The watershed algorithm is described by Vincent *et al.* [44]. The set of the catchment basins of the grayscale image I is equal to the set of $X_{h_{max}}$ obtained after the following recursion:

$$X_{h_{min}} = T_{h_{min}}(I) \tag{10}$$

where, $h_{min}$ and $h_{max}$ denote minimum and maximum gray level respectively in I. $X_{h_{min}} = \{p \in D_I, I(p) \leq h_{min}\}$ where $D_I$ is the set of values taken by the image I and $T_{h_{min}}(I)$ is the set of points which are first reached by water. These points constitute the starting set of the recursion.

$$X_{h+1} = \min_{(h+1)} \bigcup IZ_{T_{h+1}}(I)X_h, \quad , \forall h \in [h_{min}, h_{max-1}] \tag{11}$$

where,

$$IZ_A(B) = \bigcup_{i \in [1:k]} iZ_A(B_i) \tag{12}$$

$$iZ_A(B_i) = p \in A, \forall j \in [1,k] - \{i\}, d_A(p, B_i) < d_A(p, B_j) \tag{13}$$

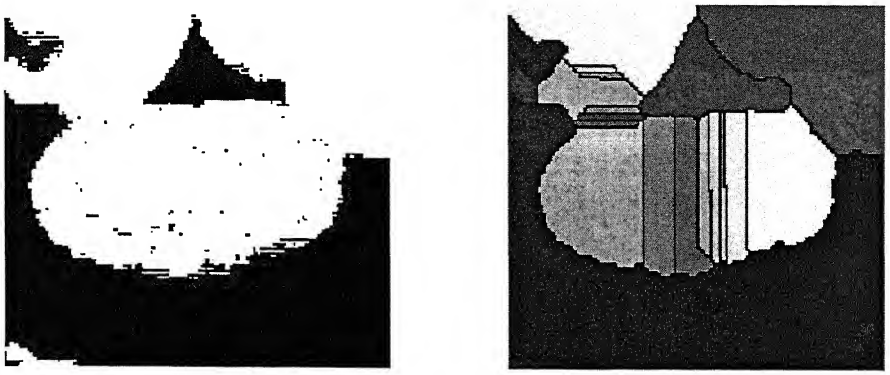


Figure 3.6: Oversegmentation : Cytoplasm mask and watershed segmented output

where,

$$IZ_A(B) = \bigcup_{i\in[1:k]} iZ_A(B_i) \quad (12)$$

$$iZ_A(B_i) = p \in A, \forall j \in [1,k] - \{i\}, d_A(p,B_i) < d_A(p,B_j) \quad (13)$$

Here, $iZ_A(B_i)$ is the geodesic influence zone of the connected component $B_i$ of $B$ in $A$, defined as the locus of the points of A whose geodesic distance to $B_i$ is smaller than their geodesic distance to any other component of $B$. *Geodesic distance* $d_A(x,y)$ is defined as the shortest path (if any) between $x$ and $y$ and totally included in region A.

We apply watershed algorithm on the binary image of the cytoplasm. As a result of declustering we obtain a cluster which belongs to the cytoplasm. The advantage of using the watershed algorithm is that the contour information is not lost. In some cases, due to oversegmentation, cytoplasm is divided into several clusters, so in order to get binary mask for cytoplasm we need to merge these clusters.

### 3.2.6 Merging

The problem with the watershed algorithm is that it is very sensitive to sharp boundaries and in effect gives rise to the cytoplasm getting divided into multiple clusters rather than single cluster as shown in figure 3.6.

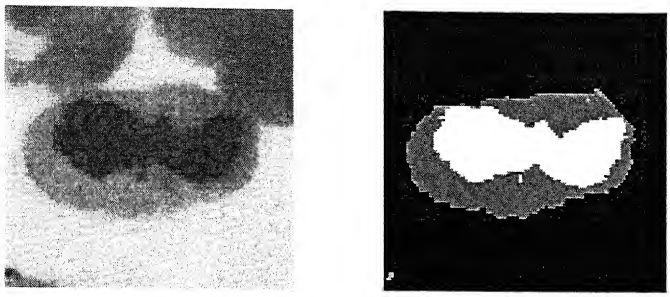


Figure 3.7: Final Segmentation

We adopt a nucleus mask based merging technique for merging clusters. Here, we assume that every cluster from the cytoplasm mask which overlaps with a segment of the nucleus mask belongs to the cytoplasm and vice-versa. Using this technique, we merge all *valid* clusters and denote it as cytoplasm mask (see figure 3.7).

## 3.3 Results

The proposed scheme has been applied on images of peripheral blood smear slides obtained by Media Lab Asia's Biomedical group at Indian Institute of Technology, Kanpur. Typical image size is 768x576. Sample input images are shown in figure 3.8. Segmentation output for various types of WBC's are shown in Figures 3.9(Lymphocyte), 3.10(Monocyte), 3.11(Eosinophil), 3.12(Neutrophil) and 3.13(Basophil). Our technique exhibits good results even in the case of varying brightness and in some cases poor contrast between cytoplasm and image background.

Although our technique produces good results for most of the cases, it is not able to identify touching cells as different and considers them as a single cell. Further, if there is dense population of granulocytes, technique can not clearly distinguish nucleus as can be seen in Figure 3.11.

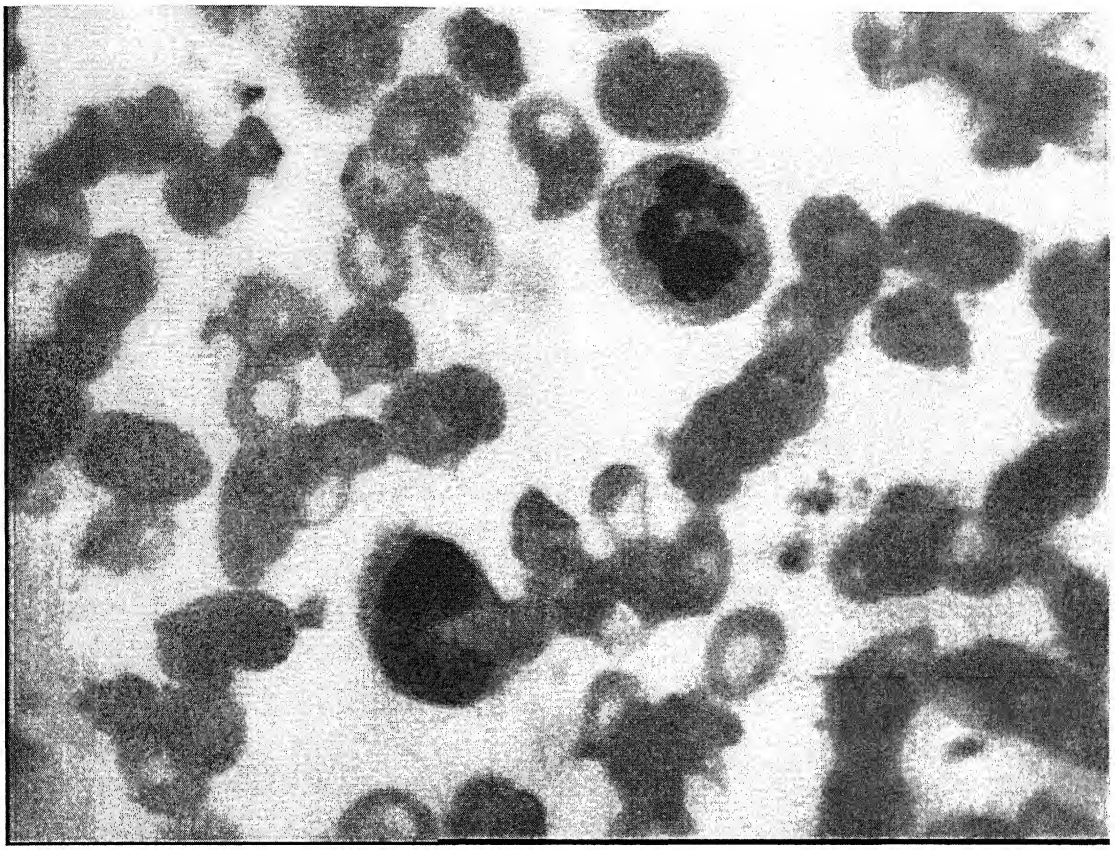


Figure 3.8: Sample input image (original size 768x576)

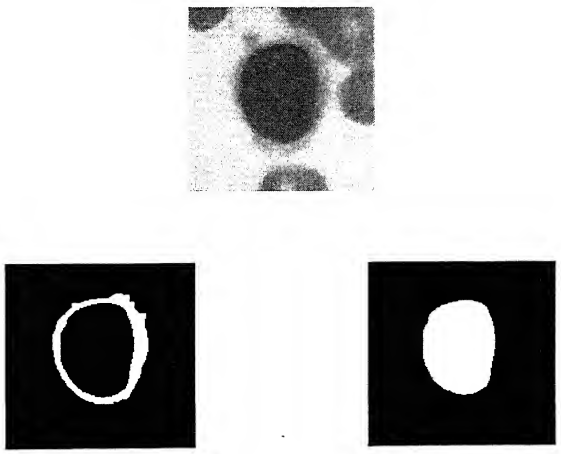


Figure 3.9: Segmented output for typical lymphocyte

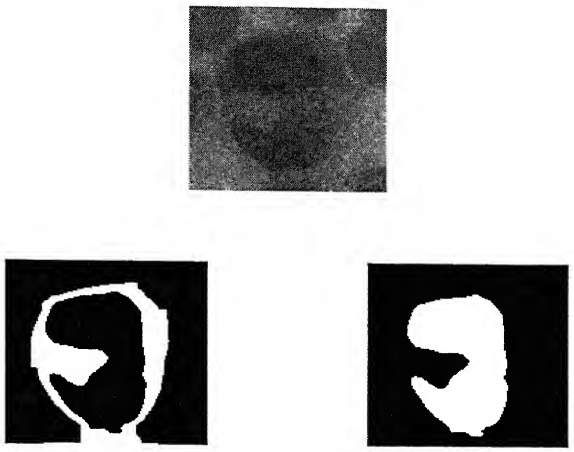


Figure 3.10: Segmented output for typical monocyte

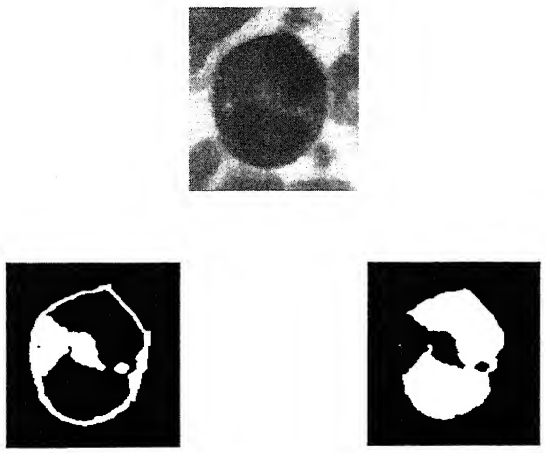


Figure 3.11: Segmented output for typical eosinophil

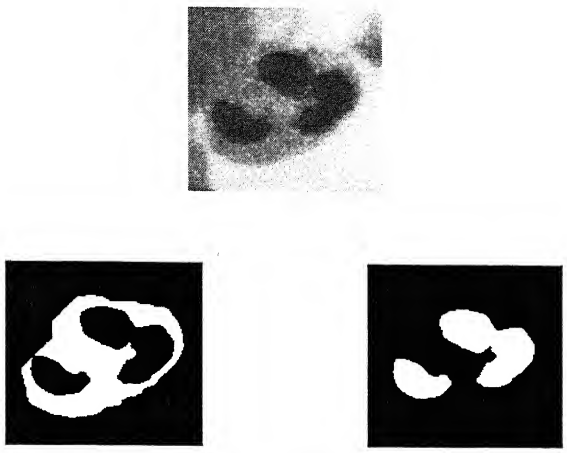


Figure 3.12: Segmented output for typical neutrophil

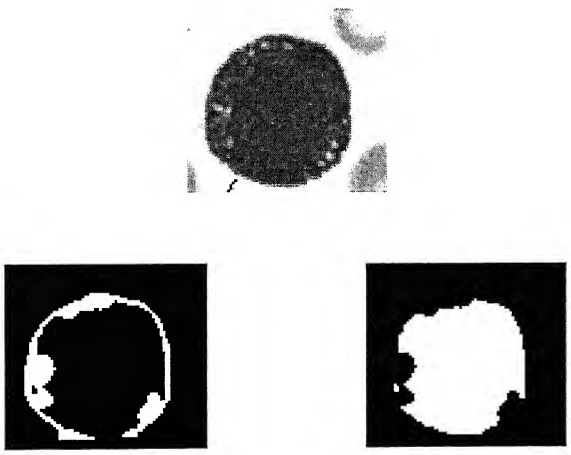


Figure 3.13: Segmented output for typical basophil

# Chapter 4

# Feature Extraction

## 4.1 Introduction

Features are representative measures of a pattern and are chosen by their ability to identify the input pattern. Feature extraction can be seen as a process of mapping the given data into useful features.

The design of a feature extractor is highly application dependent. An ideal feature extractor removes irrelevant and redundant information from data, preserving important discriminant information in order to ensure good class-separability. The advantages in working with features rather than the whole image are :

- Reduction in computational complexity of pattern classification due to reduction in dimensionality, resulting in efficient and faster classification
- Reduction in space requirement, as feature data requires much lesser space than the entire image.

## 4.2 Cell Structure

WBC's comprise of mainly two parts, the nucleus and the cytoplasm. Characteristics of these two parts vary across different types of WBC's, which enable us to classify

them as one of the five types. The quantitative features of both the cell-parts could be used to classify the cell.

White blood cells are broadly classified as Granulocytes and Non-Granulocytes, based on presence or absence of granules in the cytoplasm [7, 45]. Visually, the differences in the color, size and spread of granules serve as vital cues for distinguishing the different types of granulocytes. Non-granulocytes have a single-lobed nucleus while the granulocytes generally have multi-lobed nucleus. The cell-parts are characterized in terms of their shape and color.

### 4.2.1 Non-Granulocytes

The Non-Granulocytes are of 2 types : Lymphocyte and Monocyte

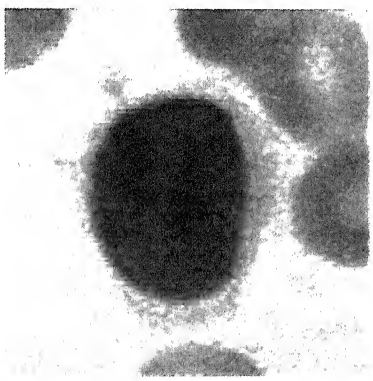


Figure 4.1: Typical Lymphocyte

*Lymphocyte* : Lymphocyte is identified by the low value of the area-ratio between the cytoplasm and the nucleus, since the cytoplasm is present only along a thin rim around the nucleus. The shape of the nucleus is generally circular, as shown in figure 4.1.

*Monocyte* : Monocyte generally shows a dent in the ellipsoidal nucleus. The shape and size of the dent are not consistent. Cytoplasm occupies fair share of total cell area unlike lymphocyte (Figure 4.2).

Figure 4.2: Typical Monocyte

### 4.2.2 Granulocytes

The granulocytes are divided in 3 classes : Eosinophils, Basophils and Neutrophils.

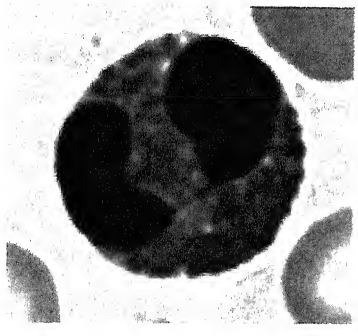


Figure 4.3: Typical Eosinophil

*Eosinophils* : Eosinophils have compactly packed, red-colored granules in the cytoplasm. The nucleus is generally bi-lobed, the lobes being linked by a ribbon-like extension. The shape of the cell boundary is generally oval (Figure 4.3).

*Neutrophils* : Neutrophils have small purple-colored granules that are loosely scattered in the cytoplasm. They have segmented nuclei with 2-5 lobes. The shape of the cell boundary is generally oval (Figure 4.4).

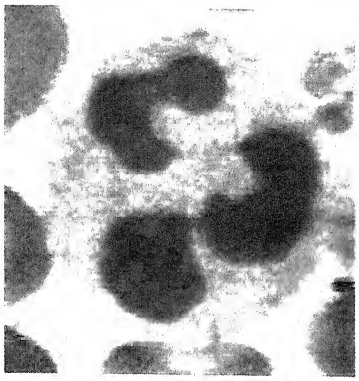


Figure 4.4: Typical Neutrophil

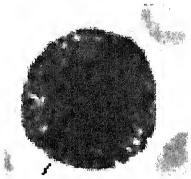


Figure 4.5: Typical Basophil

*Basophil* : Basophils have large blue-colored granules loosely scattered in the cytoplasm. They have bi-lobed nuclei often obscured by granules. The shape of the cell boundary is generally oval (Figure 4.5).

## 4.3 Cell Features

Features for discriminating between the different cell classes are based on visual cues used by experts. But converting the description precisely into metric is a difficult task. Hence, we need to experiment with different possible features and pick the best set of features for classification. We adopted the following features:

- Shape based features
- Color based features
- Texture based features

### 4.3.1 Shape based features

The shape of an object can be described using shape descriptors [18]. Shape descriptors may not be an accurate description of shape, but they must be distinct enough for different shapes for classification purpose. We use a binary mask of the cytoplasm and the nucleus to compute these features. The features used are :

*Eccentricity* : Eccentricity is defined as the ratio between the major and the minor axes. It gives a measure of how close the shape is to a circle. In our case, the eccentricity can be computed as the ratio of the eigen values of the covariance matrix of the position vectors of the foreground pixels. Let's say,

$$P_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

represent the position vectors. Then, the covariance matrix for it can be given by,

$$\mathrm{C} = E\left[(P_i - m_p)(P_i - m_p)^T\right]$$

where $m_p$ is the mean position vector. Now the eigen values are given by the values of $\lambda$ that satisfy,

$$\mathrm{C}\hat{v} = \lambda\hat{v}, \quad \text{for non zero values of } \hat{v}$$

$$Eccentricity = 1 - \frac{\lambda_2}{\lambda_1}, \quad where\ \lambda_1 \geq \lambda_2 \tag{1}$$

*Compactness* : Compactness is the ratio of the area to square of the perimeter. Area is measured as the count of the foreground pixels for a cell. Perimeter is calculated as the number of pixels lying on the boundary of a structure. Compactness is an index of the extent of indentation of the boundary. The value is high if the boundary is smooth, and low otherwise.

$$Compactness = \frac{\text{Area}}{\text{Perimeter}^2} \tag{2}$$

*Area Ratio* : Area ratio is counted as number of pixels that make up the nucleus to the number of pixels that make up the whole cell. i.e.

$$AreaRatio = \frac{\text{Pixel count of cytoplasm}}{\text{Pixel count of nucleus}} \tag{3}$$

*Number of lobes in the nucleus* : Number of lobes that make up the nucleus is one of the distinguishing features between granulocytes and non-granulocytes. However, in the case of overlapping lobes, declustering needs to be done before counting the number of lobes.

### 4.3.2 Color based features

The color features are obtained from the segmented nucleus and cytoplasm. Average value of each color component, R, G and B, of the nucleus and cytoplasm are computed.

$$Mean_C = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{4}$$

where, $N$ is total number of pixels in region of interest and $C_i$ is the corresponding color, R or G or B, component of $i^{th}$ pixel.

### 4.3.3 Texture based features

Texture is defined as a function of the spatial variation in the pixel intensities. In recent times, texture is an important image feature, especially in content-based image retrieval systems [19, 28]. In our work, we use texture features computed from the gray-level co-occurrence matrix (GLCM), proposed by Haralick [20], to quantify texture.

*Gray-level Co-occurrence Matrix (GLCM)* : Gray-level co-occurrence matrix is also known as spatial gray level dependence (SGLD). Spatial gray-level co-occurrence estimates image properties related to second-order statistics. As the name suggests, the GLCM is constructed from the image by estimating the pairwise statistics of the pixel intensity. Each element $(i, j)$ of the matrix represents an estimate of the probability that two pixels with a specified separation have gray levels $i$ and $j$. The separation is usually specified by a displacement, $d$ and an angle, $\theta$. i.e.

$$GLCM, \Phi(d, \theta) = [\widehat{f}(i, j|d, \theta)] \tag{5}$$

$\Phi(d, \theta)$ will be a square matrix of side equal to the number of gray levels in the image and will usually not be symmetric. Symmetry is often introduced by effectively adding the GLCM to it's transpose and dividing every element by 2. This renders $\Phi(d, \theta)$ and $\Phi(d, \theta + 180°)$ identical and makes the GLCM unable to detect 180° rotations.

In texture classification, instead of the individual elements of the matrix, the features derived from the matrix are used. Haralick *et al.* [20] proposed 14 features from the matrix, out of which only following are used widely.

- Contrast = $\sum_{i,j} |i - j|^2 \ \widehat{f}(i, j)$
- Correlation = $\sum_{i,j} \frac{(i-\mu i)(j-\mu_j)\hat{f}(i-j)}{\sigma_i \sigma j}$
- Energy = $\sum_{i,j} \widehat{f}(i, j)^2$
- Homogeneity = $\sum_{i,j} \frac{\hat{f}(i,j)}{1+|i-j|}$

where, $\widehat{f}(i, j)$ represents $(i, j)^{th}$ entry of GLCM i.e. the number of occurrences of the pair of gray levels $i$ and $j$ with distance $d$ and angle $\theta$ apart. $(\mu_i, \mu_j)$ and $(\sigma_i, \sigma_j)$ represent mean and standard deviation respectively of GLCM along rows

and columns and are obtained as :

$$\mu_i = \sum_i i \sum_j \widehat{f}(i,j) \tag{6}$$

$$\mu_j = \sum_j j \sum_i \widehat{f}(i,j) \tag{7}$$

$$\sigma_i = \sum_i (i-\mu_i)^2 \sum_j \widehat{f}(i,j) \tag{8}$$

$$\sigma_j = \sum_j (j-\mu_j)^2 \sum_i \widehat{f}(i,j) \tag{9}$$

## 4.4 Conclusion

For classification the following features were used :

- Area ratio
- Eccentricity of the nucleus
- Eccentricity of the cytoplasm
- Compactness of the nucleus
- Mean red value of the nucleus
- Mean green value of the nucleus
- Mean blue value of the nucleus
- Mean red value of the cytoplasm
- Mean green value of the cytoplasm
- Mean blue value of the cytoplasm
- Contrast
- Correlation
- Energy
- Homogeneity

*Discarded features* : Although number of lobes in the nucleus is an important feature, but in our case, lobes overlapped too much and were difficult to decluster, so we dropped this feature. While compactness of the nucleus is considered as a feature, compactness of the cytoplasm is discarded, because in most of the cases, the cell boundary is nearly oval and hence compactness of the cytoplasm does not serve as a distinguishing feature.

# Chapter 5

# Classification

## 5.1 Introduction

Classification is the task of assigning to the unknown test sample, a label from one of the known classes. The task for a classifier is to evaluate a given feature vector and decide the label for the vector. A good feature extractor is an essential prerequisite for a classifier, because features with good discrimination can easily be labeled using linear classifiers such as the nearest neighbour classifier, but if the patterns are very close in feature space then non-linear methods like neural networks and support vectors machines are required. The accuracy of the classifier highly depends on the quality and the amount of information that the classifier is trained with, which can be increased by providing high order feature vectors, but increasing the dimension of the feature vector may introduce increase in redundant information, which is undesirable. Hence, the feature vector is a trade-off between the permissible classification error, the complexity of classifier and the time required for classification.

## 5.2 Data used for the study

As per the rule of thumb rule for supervised learning, the number of training patterns must be 5 to 10 times the dimensionality of the feature vector. But, due to

unavailability of sufficient data, we chose a smaller training set. The training data consists of 50 samples with 10 samples from each class and the test data consists of 30 samples 5.1. The test instances used are different from the ones used for training.

| Class | Training | Testing |
|---|---|---|
| Basophil | 10 | 5 |
| Eosinophil | 10 | 4 |
| Lymphocyte | 10 | 10 |
| Monocyte | 10 | 4 |
| Neutrophil | 10 | 7 |
| Total | 50 | 30 |

Table 5.1: Data set used for the experiments

## 5.3 Supervised Learning

The aim of supervised classification is to construct a model for predicting the correct label for an unseen pattern, on the basis of the feature vector of the test pattern. A good supervised learning algorithm leads to precise labeling of unseen samples, based on the instances which are presented to the classifier along with their labels a priori. Geometry based classifiers such as neural networks and support vector machines (SVMs) rely on the estimation of the decision boundaries in actual (or higher order) feature space, making use of appropriate error-minimizing criteria. However, goodness of such a classifier relies heavily on the time-consuming training process.

### 5.3.1 Neural Networks

Neural networks are networks of non-linear computing elements, interconnected through adjustable weights. The most popular learning technique for neural networks is 'feed forward back propagation learning'. Back propagation proceeds by comparing the outputs of the network to the expected outputs, and computing an

error measure based on sum of square differences. Figure 5.1 shows a sample neural network.

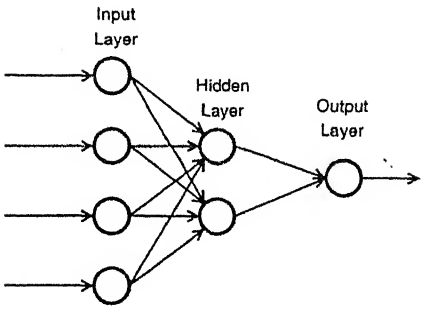



Figure 5.1: A sample neural network

### 5.3.2 Support Vector Machines

Support vector machines [8] are based on the concept of separating hyperplane. SVMs achieve classification by finding a separating hyperplane (linear or nonlinear) in a higher order mapped feature space of the data set. They are modeled as optimization problems with quadratic objective function and linear constraints. Basically, SVMs try to optimize the margin between classes. The two classes are optimally divided by a hyperplane, which does not depend on the probability distribution. It is observed that the optimal hyperplane is determined only by a small fraction of the data points, called "support vectors" (see figure 5.2). The classifier training algorithm is a procedure to find these vectors.

## 5.4 Experiments and Results

A simple feature set consisting of features based on the shape, color and texture is used for classification of WBC's. The individual performance of each category of the features is evaluated over the neural networks and the support vector machines.

We observed that using neural networks the best results are obtained using two hidden layers with 6 neurons each, when using feed forward back propagation neural

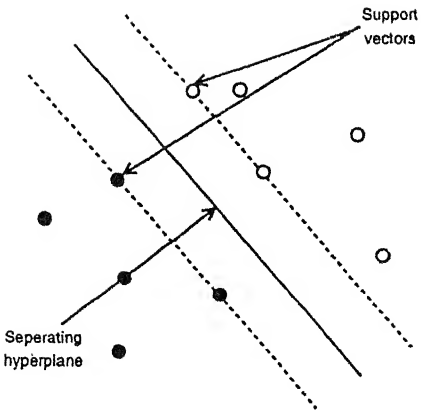



Figure 5.2: Support vectors and separating hyperplane

tworks. Number of input neurons depends on feature set we are using for the periments and number of output neurons were chosen to be 5. In the case of pport vector machines, the SVM with $4^{th}$ degree polynomial kernel yielded best sults. The performance of the combined features is also studied. Results are shown table 5.2.

| Classifier | SVM | NNet |
|---|---|---|
| Texture | 36.7 | 40.0 |
| Shape-Color | 70.0 | **80.0** |
| Combined | 70.0 | 76.7 |

able 5.2: Comparison of classifier performance on the different feature sets (in %)

We see that shape and color based features perform better than texture features. This can be explained as follows. In our case, we have a small rectangular window ontaining cells, in which texture of different classes of cells can not be captured as discriminant features. Figure 5.3 shows the distribution of 4 texture features for 50 training samples. We can observe that except for $3^{rd}$ texture feature (energy) we don't see many variations in values of features among different classes, so using only texture as features does not provide any discriminant information to train the classifier.

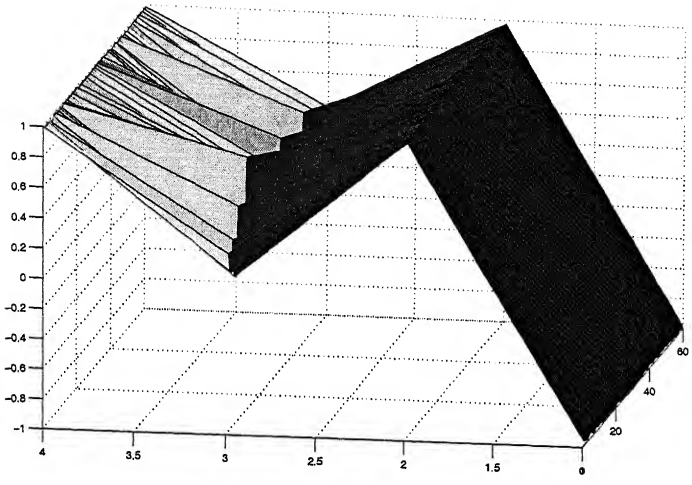



Figure 5.3: Texture features for all classes

Also, it can be seen that when texture features are combined with color and shape features, we are not getting improvement in classification accuracy, rather we observe decline in the accuracy. Including textures as features for classification probably introduces redundancy in the feature set, thus making it difficult for any classifier to mark the decision boundary between classes. Based on the results we obtained, we can say that given shape and color based feature, texture features are redundant. The classification performance is good considering limited training data and poor image quality. The best performance is obtained with the neural network classifier when trained with shape and color based features.

# Chapter 6

# Conclusion and Future Work

In this thesis, we propose an automated system to obtain the differential blood count, using image processing and machine learning techniques. The system takes as input, color images of blood smears and determines the classes of the WBC's.

We present an effective two-stage segmentation technique. At first, $k$-means is performed on the HSV-equivalent of the image in order to locate WBC's in the image. A second level of segmentation is performed using autothresholding followed by declustering using the watershed algorithm to achieve finer level of segmentation and to facilitate segmentation of the cytoplasm and the nucleus. The features chosen for classification were based on shape, color and texture of the segmented nucleus and the cytoplasm. Support vector machine and neural network classifiers were tried on different combination of feature sets of these color and shape based features using the neural network classifier has been observed to be the most effective. The peak classification accuracy of 80% was obtained using a neural network based classifier with color and shape based features.

The performance obtained is not par with the state-of-the-art works. Katz [24] obtained peak accuracy of 98% with a feature set based on cell color, size and nuclear morphological information, with the dataset containing more than 200 instances of different cells. Ongun *et al.* [30] achieved accuracy of 91% using support vector machine with 57-dimensional feature vectors which included color statistics, texture,

shape features and color histogram. Song *et al.* [42], using context-based classification, obtained accuracy of 91%, where huge training data consisted blood samples from 220 specimens (consisting of 13,200 cells). In our work, considering the paucity of data and relatively poor quality of input images, our proposed technique seems promising.

## 6.1 Future Work

- The segmentation scheme in our work is effective, but is unable to handle overlapping cells. The scheme can be enhanced by including techniques for declustering, leading to segmentation of overlapping cells as well.
- A method identifying and rejecting unsatisfactorily segmented images needs to be devised.
- Feature selection can be incorporated in order to eliminate the redundant and confusing features.
- Our work focuses on the mature classes of cells only, but can be extended for the immature classes of cells also.

# Bibliography


[1] AUNE, M. W., BECKER, J. L., BRUGNARA, C., CANFIELD, W., AND DORFMAN, D. M. Automated flow cytometric analysis of blood cells in cerebrospinal fluid : Analytic performance
*http://www.ajcp.com/previews/abstracts/203421.html.*

[2] BAMFORD, P., AND LOVELL, B. Method for accurate unsupervised cell nucleus segmentation. In *EMBS Conference* (2001), pp. 2704–2708.

[3] BAO, H. F., DEN, H. H. C., GELSEMA, E. S., AND SMEULDERS, A. W. Automated white blood cell classification revisited. *Medical Informatics 12*, 1 (1997), 23–31.

[4] BEUCHER, S. The watershed transformation applied to image segmentation
*http://cmm.ensmp.fr/~beucher/publi/pefferkorn.pdf.*

[5] BHANUPRAKASH, N., RAMAKRISHNAN, A. G., SURESH, S., AND CHOW, T. W. P. Fetal lung maturity analysis using ultrasound image features. *IEEE Trans. Inf. Tech. Biomedicine* (March 2002).

[6] BIKHET, S. F., DARWISH, A. M., TOLBA, H. A., AND SHAHEEN, S. I. Segmentation and classification of white blood cells. In *IEEE International conference on Acoustics, Speech, and Signal Processing* (June 2000), vol. 4, pp. 2259–2261.

[7] Blood
*https://courses.stu.qmul.ac.uk/smd/kb/microanatomy/blood/.*

[8] BURGES, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery 2*, 2 (1998), 121–167.

[9] CLARKE, L. P., VELTHUIZEN, R. P., CAMACHO, M. A., HEINE, J. J., VAIDYANATHAN, M., HALL, L. O., THATCHER, R. W., AND SILBIGER, M. L. MRI segmentation: Methods and applications. In *Magnetic Resonance Imaging* (1995), vol. 13, pp. 343–368.

[10] COMANICIU, D., AND MEER, P. Cell image segmentation for diagnositc pathology *http://www.caip.rutgers.edu/riul/research/papers/ps/cell.ps.gz.*

[11] COMANICIU, D., AND MEER, P. Mean shift: a robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence 24*, 5 (May 2002), 603–619.

[12] The coulter principle *http://www.beckmancoulter.com/products/instrument/partchar/technology/coulterprinciple.asp.*

[13] CSEKE, I. A fast segmentation scheme for white blood cell images. In *11th IAPR Int. Conf. on Pattern Recognition, Conf. C: Image, Speech and Signal Analysis* (1992), vol. 3, pp. 530–533.

[14] CSEKE, I., AND FAZEKAS, Z. Comments on gray-level thresholding of images using a correlation criterion. *PRL 11* (1990), 709–710.

[15] DACIE, J., AND LEWIS, S. M. *Practical Haematology.* New York Churchill Livingstone, 2001.

[16] FORSYTH, D. A., AND PONCE, J. *Computer Vision : A Modern Approach.* Pearson Education, 2003.

[17] GNIADECKA, M., PHILIPSEN, P., WESSEL, S., NIELSEN, O., CHRISTENSEN, D., HERCOGOVA, J., ROSEEN, K., THOMSEN, H., HANSEN, L., AND WULF, H. Automated malignant melanoma diagnosis: Neural network analysis of

chemical alterations presented in Raman spectra of the cancer *http://www2.imm.dtu.dk/site/publications.htm.*

[18] GONZALEZ, R. C., AND WOODS, R. E. *Digital Image Processing*, 2 ed. Prentice Hall, 2002.

[19] HARALICK, R., AND AKSOY, S. Textural features for image database retrieval. In *IEEE Workshop on Content-Based Access of Image and Video Libraries* (June 1998), pp. 45–49.

[20] HARALICK, R. M., SHANMUGAM, B., AND DINSTEIN, I. Texture features for image classification. *IEEE Trans. on Systems, Man, and Cybernetics 3*, 6 (November 1973), 610–621.

[21] HENGEN, H., SPOOR, S., AND PANDIT, M. Analysis of blood and bone marrow smears using digital image processing techniques. *SPIE Medical Imaging* (February 2002).

[22] JAIN, A. K. *Fundamentals of Digital Image Processing.* Prentice-Hall, 1989.

[23] JIANG, M. Watershed segmentation *http://ct.radiology.uiowa.edu/~jiangm/courses/dip/html/node138.html.*

[24] KATZ, A. R. J. Image analysis and supervised learning in the automated differentiation of white blood cells from microscopic images. Master's thesis, RMIT, 2000.

[25] KOVALEV, V. A., GRIGORIEV, A. Y., AND AHN, H.-S. Robust recognition of white blood cell images. In *Proceedings of the 13th Int. Conf. on Pattern Recognition* (August 1996), vol. 4, pp. 371–375.

[26] LIN, W., XIAO, J., AND MICHELI-TZANAKOU, E. A computational intelligence system for cell classification. In *IEEE International Conference on Information Technology Applications in Biomedicine* (May 1998), pp. 105–109.

[27] LIU, Z., LIEW, H., CLEMENT, J., AND THOMAS, C. Bone image segmentation. *IEEE Trans. on Biomedical Engineering 46*, 5 (May 1999), 565–573.

[28] MANJUNATH, B., AND MA, W. Texture features browsing and retrieval of image data. *IEEE Trans. on Pattern Analysis and Machine Intelligence 8*, 18 (1996), 837–842.

[29] Medical encyclopedia *http://www.nlm.nih.gov/medlineplus/ency/article/001299.htm.*

[30] ONGUN, G., HALICI, U., LEBLEBICIOGLU, K., AND ATALAY, V. Feature extraction and classification of blood cells for an automated differential blood count system. In *International Joint conference on Neural Networks* (July 2001), vol. 4, pp. 2461–2466.

[31] OTSU, N. A threshold selection method from gray level histograms. *IEEE Trans. on Systems, Man and Cybernetics 9* (March 1979), 62–66.

[32] PARK, J.-S., AND KELLER, J. M. Fuzzy patch label relaxation in bone marrow cell segmentation. In *IEEE International Conference on Computational Cybernetics and Simulation* (1997), vol. 2, pp. 1133–1138.

[33] PITAS, I. *Digital Image Processing Algorithms.* Prentice-Hall, 1993.

[34] PRAKASH, K. B., AND RAMAKRISHNAN, A. MR image enhancement by non-linear techniques. In *National Conf. on Communication* (January 2002), Indian Institute of Technology, Bombay.

[35] RACHID, S., NIKI, N., NISHITANI, H., NAKAMURA, S., AND MORI, S. Segmentation of sputum color image for lung cancer diagnosis based on neural networks. In *ICIAP (2)* (1997), pp. 461–468.

[36] REBAR, A., MACWILLIAMS, P., FELDMAN, B., METZGER, F., POLLOCK, R., AND ROCHE, J. *Laboratory Methods in Hematology.* International Veterinary Information Service, 2004.

[37] REDDI, S., RUDIN, S., AND KESHAVAN, H. An optimal multiple threshold scheme for image segmentation. *IEEE Trans. on Systems, Man and Cybernetics 14*, 4 (July 1984), 661–665.

[38] Segment blood vessels from 2-d digital angiogram *http://www.inf.u-szeged.hu/~ssip/2001/projects/documentation/project22.*

[39] SHEIKH, H., ZHU, B., AND MICHELI-TZANAKOU, E. Blood cell identification using neural networks. *Proceedings of the 1996 IEEE Twenty-Second Annual Northeast Bioengineering Conference 21*, 4 (1999), 119–120.

[40] SINHA, N. Segmentation and classification of colour images of blood cells. Master's thesis, Indian Institute of Science, 2003.

[41] SOBREVILLA, P., MONTSENY, E., AND KELLER, J. White blood cell detection in bone marrow images. In *Proceedings of the 18th NAFIPS International Conference* (June 1999).

[42] SONG, X., ABU-MOSTAFA, Y., SILL, J., AND KASDAN, H. Incorporating contextual information in white blood cell identification. In *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10* (Cambridge, MA, USA, 1998), MIT Press, pp. 950–956.

[43] UNNIKRISHNAN, K. P., AND VENUGOPAL, K. P. Alopex: A correlation-based learning algorithm for feedforward and recurrent neural networks. *Neural Computation 6*, 3 (1994), 469–490.

[44] VICENT, L., AND SOILLE, P. Watersheds in digital spaces: An efficient algorithm based on immersion simulations. *IEEE PAMI, 1991 13*, 6 (1991), 583–598.

[45] WERMSER, D., HAUSSMANN, G., AND LIEDTKE, C. Segmentation of blood smears by hierarchical thresholding. *CVGIP 25*, 2 (February 1984), 151–168.

[46] White blood cells *http://en.wikipedia.org/wiki/white_blood_cell.*

[47] ZUCKER-FRANKLIN, D. *Atlas of Blood Cells.* Lea & Febiger, Philadephia,PA, 1981.